Final Report

of the

Database Standardization Committee

June 8, 1998

Prepared for
the
National Defense Industrial Association
and the
Department of Defense Modeling and
Simulation Community

TABLE OF CONTENTS

LIST OF FIGURES

FOREWORD

Science Applications International Corporation (SAIC) and Presearch Incorporated are pleased to consolidate the efforts of the National Defense Industrial Association (NDIA) Database Standardization Committee. A complete listing of the individuals who participated in this study is included in Section 2 - Study Organization. Their organizations and corporations are shown below. SAIC and Presearch Incorporated look forward to the savings in time and budget dollars to the Defense community, for both the government and the private sector, which this effort should afford.

National Security Industrial Association (NSIA)

National Defense Industrial Association (NDIA)

HQ USAF, Directorate of Modeling, Simulation, and Analysis

HQ USAF, Directorate of Command and Control

Air Force Studies and Analysis Agency

Northrop Grumman Corporation

Aeronautical Systems Center, FBI and XRA

Raytheon Systems Company

Lockheed Martin, Inc.

Electronic Systems Center, Modeling and Simulation Division

Science Applications International Corporation

Boeing Company

TRW/Ballistic Missile Defense Office (BMDO)

Presearch Incorporated

SPAWAR 131

Lockheed Martin Tactical Aircraft Systems

SECTION 1 — INTRODUCTION

EXECUTIVE SUMMARY

Introduction. The National Defense Industrial Association (NDIA) Database Standardization Committee was formed in August 1996. Its purpose was to review existing database issues and programs in the Department of Defense (DoD) modeling and simulation (M&S) community and recommend improvements to the study process.

Approach. The ad hoc committee was chartered originally by the National Security Industrial Association (NSIA), which evolved into the American Defense Preparedness Association/NSIA, and finally into the NDIA. The committee members were drawn from both the DoD military services and the private sector supporting DoD for M&S to support acquisition programs. A balance was sought between government, industry primes, industry avionics, and industry study houses.

The basic premise was that, unlike model management, no direction existed for coherent, consistent, documented system data. It was observed that quick-reaction study response is difficult when large databases must be rebuilt, and too much time and money is wasted on initial database efforts prior to conducting a study.

The committee began their efforts by outlining the following objectives, which, when thoroughly investigated, would lead to a briefing to government and a final report which would be available to government and industry through NDIA. The objectives selected were to:

1. Identify data sources.
2. Recommend standards and processes.
3. Review data sharing methods.
4. Identify advantages and disadvantages (of sharing data).
5. Recommend ways to avoid misuse.
6. Prioritize models needing databases.
7. Propose an implementation plan.

Conclusions. Through the exploration of the objectives, a major void in the database support to the study process was identified. Databases were not maintained and updated for reuse, as was true for many of the major models. It was found that the Air Force Pedigree process satisfied the requirement for coherent, consistent, documented data.

The study concluded it would be cost effective for the Government to adopt a process to develop and maintain databases with known pedigree to support a preferred set of standardized models.

This process would not only ensure greater consistency and credibility among studies, but would allow the studies to begin sooner and be conducted faster and more economically than the process allows today. Both government and industry would share in the cost avoidance.

Recommendations. The committee recommends that the first step is to write a policy for Pedigreed Database use and management. Once the policy is in place, the next step is to prioritize the models used most frequently and apply the Pedigree process to the databases for those models. To do this, the group recommends sufficient funding to initiate and sustain the process.

Part of the process must include certification to provide credibility and prevent misuse. The committee recommends AF/XOC (HQ USAF, Director, Command and Control, Deputy Chief of Staff/Air and Space Operations) as the implementation sponsor and Pedigree process owner of the initiative that originated at the Air Force Materiel Command (AFMC). This study should be shared with the other services and the Defense Modeling and Simulation Office (DMSO).

Following policy implementation, NDIA should investigate the technical aspects of the implementation. Finally, the group recommends this report be made available through NDIA and that NDIA assist in placing it "on-line" with the Defense Technical Information Committee (DTIC).

SECTION 2 — STUDY ORGANIZATION

STUDY PARTICIPANTS

The study participants were drawn from both the DoD and its supporting private sector. To ensure a broad perspective, volunteers were solicited from all the Services, manufacturers of weapons system delivery platforms, weapons, and avionics, and study companies.

The members of the study elected a chairman, Mr. Vinton Cline, to facilitate the efforts, and consultants from the AFMC and Headquarters Air Force were invited to participate and provide detailed reports of previous activities related to the Pedigree process or policy plans.

The following individuals responded and participated in the committee's activities:

Name	Organization or Company
General Thomas C. Richards, USAF (Ret) Committee Advisor	President National Security Industrial Association
Lt. General Lawrence F. Skibbie, USA (Ret) Committee Advisor	President National Security Industrial Association
Maj. General Thomas R. Case, USAF Committee Advisor	HQ USAF, Former Director of Modeling, Simulation, and Analysis
Maj. General Charles R. Henderson, USAF Committee Advisor	HQ USAF, Former Director of Command and Control
Brig. General James E. Sandstrom, USAF Committee Advisor	HQ USAF, Director of Command and Control
Colonel Thomas L. Allen, USAF Member	Commander, Air Force Studies and Analysis Agency
Mr. Kent Attridge Member	Northrop Grumman Corporation
Mr. Larry Beasley Consultant	Aeronautical Systems Center, (ASC/FB) Wright-Patterson AFB, OH
Mr. Peter Carellas Member	National Defense Industrial Association
Ms. Nancy Clements Consultant	Aeronautical Systems Center, (ASC/XR) Wright-Patterson AFB, OH
Mr. Vinton Cline Member, Chairman	Raytheon Systems, Inc.

Name	Organization or Company
Mr. Mort Forker Member	Lockheed Martin, Inc.
Colonel Hoot B. Gibson Member	Chief Modeling, Simulation Hanscom AFB, MA
Mr. Larry Janning Member	Science Applications International Corp.
Ms. Myong Suk "Tea" Kim Consultant	HQ USAF, Directorate of Modeling, Simulation, and Analysis
Dr. Arthur Eugene Olson Member	Boeing Company
Mr. John Phillips Member	TRW / BMDO
Mr. R. Scott Saunders Member	Presearch Incorporated
Mr. Chris Peace (deceased) Member	SPAWAR 131
Dr. Bud Simrin Member	Lockheed Martin Tactical Aircraft Systems
Mr. Pete Smith Member	Presearch Incorporated

STUDY ACTIVITIES

Date	Activity	Location
August 19 and 20, 1996	Initial Meeting	Fairfax, Virginia
September 26 and 27, 1996	2nd Meeting	Dayton, Ohio
October 30 and 31, 1996	3rd Meeting	Plano, Texas
December 2 and 3, 1996	4th Meeting	Orlando, Florida
February 3, 4, and 5, 1997	5th Meeting	Los Angeles, California
March 27, 1997	Brief to ADPA/NSIA	Arlington, Virginia
May 5, 1997	6th Meeting Brief to AF/XOC	Arlington, Virginia The Pentagon
May 29, 1997	Briefing to Navy (N81)	The Pentagon

SECTION 3 — THE STUDY

DATABASE STANDARDIZATION

Standardizing databases for reuse requires the development of consistent, coherent, and documented data. In a sense, these are the same factors that have long been applied to the development and management of models used by military analysts. The analytic community has developed a full range of models, from engineering through the campaign level, which are documented and updated regularly to take advantage of lessons learned and new information about emerging concepts, tactics, and technology.

While the models and analysts who use them have benefited from this process, there has been a void in the application of similar procedures to databases. All too often, when a study is chartered, databases must be regenerated. This delays the study start and increases the study cost because of the requirement to find and generate the data and costs associated with keeping a study team together for longer periods.

Concise, coherent, documented, and readily available databases can be quickly adapted for use by others, resulting in a more efficient, lower-cost study process. Of course, to do this, many factors must be considered. Standards must be set and guidelines for sharing and access must be determined. Priorities must be outlined for the most frequently used databases to receive the greatest attention. A technical implementation plan must be set into motion once the policy is set.

DATA SOURCES

The goals of the Data Sources objective were to identify, by source, where data currently resides, to suggest where it might reside in the future, and to recommend how it might be managed.

The first task undertaken to support this objective was to identify the various data types currently used to support modeling, simulation, and analysis. The following data types were identified:

- Threat System Data
- Blue System Data/Equipment Data
- Force Descriptions, Concepts of Operations/Joint ConOps
- Scenario Data/Order of Battle
- Environmental Data
- Test Results and Analysis Data
- Human Factors

The next task was to identify where this data is located. This focused primarily on identifying the originator or maintainer of these data types. The following list is not complete, but represents the types of information observed:

- Threat Data - Defense Intelligence Agency (DIA)
- Blue Systems - Program Offices
- Force Descriptions ConOps - Major Commands
- Joint ConOps - Joint Publications and Theater CINCS
- Scenario Data/Order of Battle - DIA, Analysis Centers
- Environmental Data - AWS, Phillips Lab, . . .
- Test Results and Analysis Data - Ranges, Labs, System Program Offices (SPOs)
- Human Factors - Armstrong Lab

Figure 1 illustrates where many of the databases currently reside and how they interact.

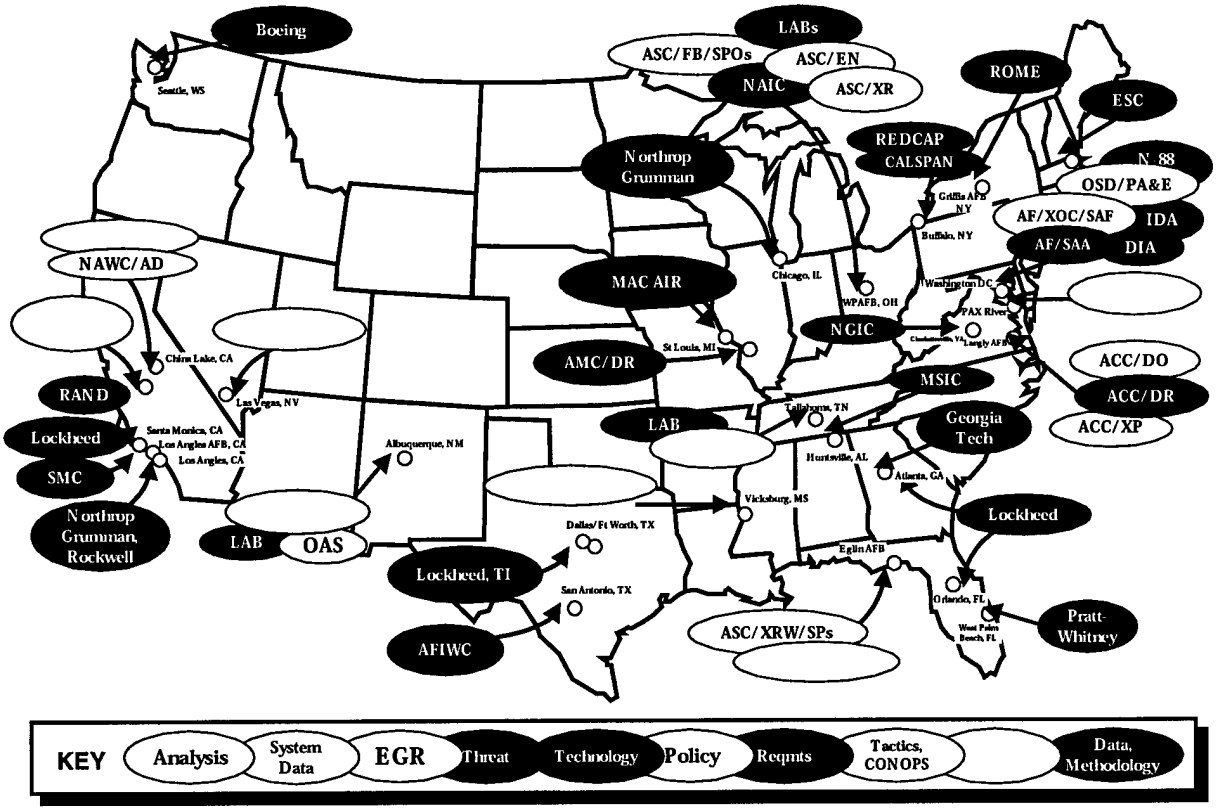


Figure 1. Current Pedigree Database Interactions

As can be seen, data to support modeling, simulation, and analysis currently resides in numerous locations. The next step in this investigation determined if anyone was centralizing a repository for this data. A brief search on the worldwide web revealed the DMSO had instituted a working group specifically targeting at maintaining a data repository. The Authoritative Data Sources (ADSs) group had formed with short-term and longer-term missions.

The short-term missions were to:

- Identify, describe, and provide Data Sources, ADSs, and Data Warehouses on the Modeling and Simulation Resource Repository (MSRR).
- Prioritize M&S community data requirements and make their descriptions available on the MSRR.

A long-term mission was to:

- Provide required, formatted instance data in a timely manner.

The group's overall purposes were to:

- Identify and categorize DoD data sources and data interchange formats (DIFs) to support the M&S community.
- Obtain Component Identification/Approval of ADSs.
- Expand and describe types of data sources from existing databases to include other types of sources (e.g., official documents and DIFs).
- Prioritize (based on user needs) the categories of data sources and existing DIFs to be identified.
- Develop, populate, and publicize a data source directory on the MSRR.
- Develop and continually expand an ADS taxonomy with definitions.
- Produce and maintain an ADS data model.
- Define data centers, data sources, ADSs, customers, and their responsibilities.
- Obtain component approval of definitions and responsibilities.
- Identify customers.
- Provide out-year's requirements.

The group went on to further define their terms and responsibilities:

- DATA SOURCE
  - An organization or subject matter expert who, because of either mission or expertise, serves as a data producer.
  - Responsible for providing the data, properly safeguarding data in accordance with (IAW) established security regulations and documenting, at a minimum, the accuracy, completeness, and currency of its data. Data sources should be able to describe, in detail, their data and/or any application of system utilized in creating and/or maintaining that data.
- AUTHORITATIVE SOURCE
  - A data source whose products have undergone producer data Verification, Validation, and Certification (VV&C) activities.
  - In addition to the responsibilities of a data source, an ADS is responsible for ensuring producer data VV&C activities have been accomplished.
- DATA CUSTOMER
  - An organization that uses data from a data source or center.
  - Responsible for (1) using the data according to any agreements with data sources and centers, (2) conducting user data VV&C activities, and (3) properly safeguarding data IAW established security regulations. Customers will report any suspected problems with the data to the supplying data source or data center as appropriate.
- DATA CENTER
  - An organization serving as a conduit between data sources and data customers. The data center may transform these data as necessary to meet the operational requirements, format, security, and data VV&C provisions of its sources and supported users.
  - Responsible for: (1) collecting and providing data to customers IAW arrangements with data sources; (2) safeguarding data IAW established security regulations; and (3) ensuring correct transformation of data between the source and the customer.
  - In addition to the instance data, the data center should store metadata that, at a minimum, consists of information on the data source and point of contact (POC); creation date; VV&C information; other users of this data; and caveats.

As a result of the work done thus far, DMSO developed a taxonomy of ADSs. This list is as follows:

1. Doctrine and Operations
2. Environment
3. Scenario
4. Force Description
5. Equipment
6. Service Support
7. Unit Performance
8. Test Results
9. Human Factors
10. Political
11. Finance
12. Metadata/Standards
13. Miscellaneous

Based on this taxonomy and the work done by DMSO to date, it was decided that ADSs either do, or someday will, exist to support modeling, simulation, and analysis. These sources, however, either provide engineering and physics-based descriptions of systems and environments, or maintain specific data formatted for use in a specific pre-identified model or simulation. The sources do not typically provide a data repository for all models that might be used in a general study area.

To meet this need, the Air Force should adopt a Pedigree process to support an identified analytic toolkit of models and simulations for its major study areas. This process should be supported through existing model management mechanisms or through a centralized coordination function designated by the Air Force. The result would be ADSs for studies within the Air Force, which could link the Pedigree process into the DMSO taxonomy.

CERTIFICATION STANDARDS

Key to the Pedigree process is the acceptance and application of data standards by all data sources supporting the process. These standards range from those implementing specific technical formats and descriptive requirements, to the full range of guidelines for documentation, definition, derivation, consistency, and transparency. These latter standards are particularly important for derived data resulting from interactions of systems and models, such as vulnerabilities or probabilities of kill. Physics-based descriptions, such as size, color, electrical power, and speed must be tied to a reference base through documentation. Regardless of the specific information, certified data must be provided in context (scenarios, environmental conditions, etc.) in order for data managers and analysts to apply the data in appropriate ways. Specific standards necessary to support various levels of certification are provided next.

Technical Standards. Data used to support Air Force models and simulations must meet, at the most basic level, the format and technical requirements established by the Institute of Electrical and Electronic Engineers (IEEE), the group responsible for evolving international data standards

for all users of computers, computer models, and information technology. In addition, Air Force data must meet the technical, format, and characterization guidelines being developed within the Department of Defense by DMSO and by the Joint Data Support System. The Joint Data Support System is the single point of contact under the Director, Programs, Analysis, and Evaluation, for supporting the Joint Analytical Model Improvement Program.

As these standards are developed, refined, and promulgated, Air Force data and model managers will need to interact with these organizations, both to help shape the standards so they meet Air Force capabilities and to ensure Air Force data is captured and represented in accordance with these standards.

Documentation. Data, like models, must be well documented and configuration controlled. Documentation must accompany all applications of the data, so analysts, users, and supported decision makers will know where the data came from, when it was developed, and what models and methodologies were used to produce derived data. In addition, they must know what changes to source data were made to fit specific study applications.

A major debate in evaluating the results of any study involves the source and credibility of the data used to run the models. Documenting the data source or pedigree can limit the debate and help analysts and decision-makers determine whether other data would be more appropriate for the application. Without documentation, the data Pedigree is unknown, and no final judgment can be rendered, either on the data validity, or on the credibility of the study results upon which the analysis rested. Full documentation should include specific dates and sources of data (tests, historical research, etc), and for derived data, the source of inputs, date of runs, and model versions, assumptions, and scenarios used to create the data.

Derivation Process. While it is necessary to document the process that produces data, documentation alone is not sufficient to ensure credibility. For example, there are any numbers of models that can be run to generate effectiveness tables for surface to air missiles (SAMs) engaging a variety of penetrating aircraft. Documenting a process that applies a spreadsheet methodology developed by a non-expert in this field can help analysts and decision-makers understand the usefulness of data derived from that source, but will not ensure credibility of a study product using this data.

There are a number of well recognized scientific models and methods available to generate data endorsed by such groups as the Joint Technical Coordinating Groups for Air Survivability (JTCG/AS). The degree to which the Pedigree process is carried to would vary depending upon the study size and criticality. The highest levels of credibility would be assigned to data that is derived from analyses involving Pedigreed models utilizing Pedigreed databases.

Consistency. Data credibility also requires consistency across the broad range of standard or approved models and simulations. Different, scientifically-supportable methods to investigate a phenomenology should provide consistent results or else the data cannot be fully pedigreed. For example, if two methods, which enjoy equal credibility in the laser range finder distance

estimation arena, provide dramatically different estimates of the effective range of a candidate system, the information from both methods must be captured and documented, along with the different estimates. Analysts who are aware of this discrepancy can design their studies to explore the sensitivity of their output to these differences and so inform other users and decision-makers as to the implications of this problem.

As science evolves and one technique is determined to be more accurate than the other, data from the less preferred methodology can be dropped and the more accepted method's results used. Data from the preferred estimating methodologies can in turn be replaced by test data or data from operational experience to ensure consistent data and a consistent understanding of performance is used in all models and simulations using the candidate systems.

In the higher level models, attributes of systems must have consistent impacts on the outcomes of interactions. Pedigree data sources must make clear to data users the consistency of data and how inconsistencies can be captured and tracked to provide the most robust study efforts.

Update Process. As earlier paragraphs indicate, data, and particularly derived data, is not a static resource. As scientific understanding of the environment and phenomenology evolves and as new systems, countermeasures, and counter-countermeasures are developed and fielded, the data associated with military effects and warfighting will also change. At one end of the spectrum, a new understanding of a revolutionary enemy capability could alter results in every level of military modeling, from engineering through campaign. A previously unknown capability could have sweeping impacts on the outcome of individual one-on-one engagements, potentially changing battle outcomes, and thereby becoming the dominate variable in studies of certain scenarios and timeframes.

Clearly, changes such as these do require a complete update of the derived databases, so that ongoing studies do not omit a major factor in their analyses. On the other hand, most changes are not revolutionary, but rather reflect improvements at the margin (a ten-percent increase in power, a five-percent improvement in radar tracking capability, an eight-percent reduction in radar reflectivity, etc). This may be important for studies in a narrow area, but it will not have a significant impact on studies across the broad array of military activity. The government needs to develop a process by which study data can be evolved through time, possibly providing an annual update schedule, but with procedures to handle changes of major significance out of cycle.

At the same time, the government needs to provide a means by which data development efforts by one group can be communicated throughout the community, so that if other groups are interested, they can join in the effort, rather than duplicate the work. By the same token, all data development efforts should be monitored and new derived data eventually made available, with documentation of source and methods to the larger community.

Supporting Models and Methods. While the focus of this committee's effort has been on data, the importance of an accepted hierarchy of models and methods for both data generation and analysis was reinforced at every meeting. For data sources, such as program offices, it is difficult

to provide an accurate characterization of a produced system to the level of engineering detail necessary to explore specific phenomenology in a laboratory development environment. However, providing derived data to support engagement, mission, and campaign-level modeling is even more challenging, particularly since dozens of models require the performance and effectiveness data. In fact, one survey estimated there are over 200 different representations of the F-15 currently resident in military models. Few of these capture F-15 performance and effectiveness in the same way. Because of the myriad of model developers and model managers, there is no ability for the F-15 Program Office to ensure that new understandings or added performance features of the F-15 are appropriately updated in all representations in all models.

Next generation modeling capability will include JSIMS (Joint Simulation System), JWARS (Joint Warfare Simulation) and JMASS (Joint Modeling and Simulation System). This joint triad may make it possible to limit the number of representations of systems, while the emerging policy on simulation-based acquisition and requirements for program offices to develop and field digital system models will help bring coherency to this arena.

In the meantime, data requirements and consistency goals can best be met by Air Force efforts to standardize its legacy model toolkit and limit its supported models to those with the greatest usefulness and credibility in the field. The Air Force can develop databases associated with and supporting these tools, by endorsing an accepted hierarchy of models and methodologies and, as a first step, always trying to update and apply these tools before supporting the development of new models. This reduces the costs and confusion associated with different groups throughout the military who maintain separate and conflicting database representations of system capabilities.

Process Enhancements. A number of actions and tools can be initiated to support the development of a Pedigree data process. In particular, the same discipline applied to model development must be applied to database development. Databases need to be monitored by users' groups, with updates managed by configuration control groups following established standardized processes, security guidelines, and library processes. Because maintaining a Pedigree database for a preferred set of models will help all users, both in the government and in industry, rules must be established to enable industry to be part of the Pedigree process.

A full array of companies would be interested in participating in the update process, particularly if they could use the resulting updated databases. This means that current security and proprietary data rules will need close examination so the benefits of a Pedigree process can be achieved without violating the protection of highly classified or proprietary data. At the very least, common scenario and defense planning guidance information needs to be promulgated, along with effects-based data that does not compromise the specifics of how that data is generated.

DATA SHARING

Data sharing among the government and contractor M&S users is the key feature of attaining highly reliable M&S results at the most economical costs to the community overall. Distributed/ shared M&S studies and analyses efforts will provide truly cost-effective end products in the same fashion that distributed wargaming and training evolutions provide the most cost effective methodology by allowing the participants to function in their real-world environments. Factors to be considered in the design of and data-sharing concept include:

- Classification and proprietary nature of data (releasability).
- System capacity requirements for bandwidth, transmission speed, latency, etc.
- Networking requirements for simultaneous usage/sharing of data during model application.
- Multi-level security issues as relate to varied levels of user community access.
- Data administration functional requirements in coordination with data developers, verification, validation and accreditation/certification of data.
- Migration of data to emerging joint systems (JSIMS, JMASS, JWARS, etc.).

Modeling and Simulation Data Administration Strategic Plan (M&S DASP) The DMSO issued the M&S DASP in April 1995. Included in its objectives is the provision of a "distributed repository system" to facilitate developer and end-user access to M&S information resources.

Since implementation, the M&S DASP program has a number of accomplishments to its credit including:

- Developed the MSRR.
- Developed data standards.
- Provided a distributed repository system to include on-line directories.
- Defined specific M&S security requirements for access across repositories.
- Developed a baseline version of the M&S taxonomy.
- Provided a M&S data administration roadmap, broken down into these time frames:
  - Near-term ('95-'96)
  - Mid-term ('97-'98)
  - Long-term (2000+)

ADVANTAGES AND DISADVANTAGES OF SHARING

As with most processes and toolsets, there is a spectrum of opinions on which approach is best, and why. While there are both pros and cons for having a data sharing process, there are overwhelming benefits to be gained in terms of cost and manpower savings in having one. However, because no set of models can be suitable for all purposes, there will always be instances where non-Pedigreed approaches must be taken. Having a process for a common set of databases has many advantages and does not preclude taking that different approach when necessary.

Advantages. The advantages can best be summarized as reduced program cost, increased ability of programs to verify results in order to make correct decisions, and ever-increasing quality of the M&S toolset. In the past, each program, company, or government office would build its own analytical databases. This resulted in dozens of organizations duplicating efforts, building similar databases for use with the same model suite. When a study house got tasked to execute a major study, a large portion of their expense was in putting together the analytical infrastructure (models and databases) to execute the study. The results would get documented, but the database development was ignored, and the whole process would be started over for the next study. When the customers then went forward to defend a study at a higher level, a large portion of the time would be spent in discussing the infrastructure, ground rules, and assumptions, and not in fully investigating all the results. A set of standard Pedigreed databases offers the following advantages:

Less duplication: Multiple organizations can use the same databases, ground rules, and assumptions rather than creating their own, saving time and money.

Reuse: Succeeding studies related to a previous study can begin with the previous set of databases and make improvements instead of starting over.

Consistent ground rules and assumptions (G&A): Studies often look 10 to 15 years into the future, and because of that, every M&S toolset requires a certain number of assumptions about how future warfare will be conducted. Decision makers at the highest levels would like to have one evolving set of G&A they approve one time and then modify as necessary for succeeding studies versus recreating every time.

Consistent data and decisions: Today different studies each develop their own databases even when they may use the same set of Pedigreed models. With a Pedigreed database process in effect, they would use the same baseline data as well as G&A. This would result in consistent study outputs, eliminating many of the discontinuities that occur today. Because the basis for making decisions by different programs can now be based on consistent results, better and more consistent decisions will be possible.

Increasing credibility: Customers at all levels are more apt to put increasing credibility into a modeling toolset if the data comes from the real source authorities, and a previously approved toolset is used as the foundation for the next study.

Easier to get started: Analysts do not have to start over each time if succeeding studies begin with previously approved ground rules, assumptions, databases, and other modeling conventions, and then continue to work out the next set of problems or bugs in the analytical processes. Programs now have a well-defined starting point available if desired.

Increasing quality of data: To coordinate the applicable databases with all the respective data source authorities as depicted in Figure 1 each time an organization undertakes a study is an impossible task; time and resources normally do not allow it. That is why a large portion of databases are constructed in-house with best-guess data by those executing the study. But over time, continued reuse of previous databases allows for an ever-increasing number to be approved by the source authorities in an organized manner.

Synergy of studies: Studies can be integrated together when needed models and databases are the same or similar.

Coordination easier: Coordinating the next study with all the applicable agencies is much easier and much quicker if the starting point was a toolset that had been previously reviewed and approved.

Model users groups must coordinate: In general, higher level models are dependent on roll-up data from lower level models. This means various model managers and model users groups must now coordinate to ensure the data outputs and formats from one model can meet the data and format needs of the next model.

More difficult for third party to challenge results: As a toolset is used, reviewed, and approved time and again by senior leaders, it becomes much more difficult for a third party to challenge the results, so senior leaders have an easier time of responding to criticism.

In summary:

Reduced cost to government: All the above advantages result in reduced cost for future studies. Typically 60 to 80 percent of the study's time and cost is involved in putting together the entire toolset/infrastructure needed to conduct the study. Having a set of documented and approved databases for reuse can cut the cost of a study from 50 to 75 percent. A major Analysis of Alternatives (AOA) that currently costs $2M to $4M can be done for under a million dollars, as was the case with the recent Joint Air-to-Surface Standoff Missile (JASSM) AOA versus the nearly $4M spent on the Advanced Medium Range Air-to-Air Missle (AMRAAM). A smaller study like the F-15 Electronic Countermeasure (ECM) AOA was accomplished for $400K to $500K versus the previous $1.5M.

Reduced cost to industry: Once a set of models and databases are documented, they can be distributed and reused. Companies who then agree to use the government-provided toolset would save significant dollars rather than developing and maintaining a unique M&S toolset in-house. This in turn will save programs money.

Disadvantages. There is also a downside to having a common set of Pedigreed databases the majority of the M&S community uses. The major disadvantages are data misuse and the administrative burden to maintain the toolsets. If senior management maintains a good awareness of the possible downside, they will be prepared to overcome any potential problems. Areas to consider include:

Administrative resources/funding required: It will take government resources to administer a Pedigreed database process, but this will probably be small in comparison to the savings. Most of the work to update databases will be done as part of the study process paid for by customers/ programs. Only the actual documentation and data coordination will require separate funding. The amount of funding will be dependent on how many models the government wants to carry in their M&S toolset, as well as how complicated the model is. The administrative cost would likely vary from $50K for a small model to $1.0 million for a large and complex model.

Users groups' workload: The database process must be coupled with the model management process, thus complicating the task of the users groups. In some cases, users groups must be formed where they do not exist and this requires additional man-hours. The bigger the user groups the more complex and expensive the coordination effort. Probably the best way to control the size of users groups and ensure only those organizations with a real vested interest participate is to charge user fees of $10K to $15K per group. This would include access to and distribution of both the model and accompanying databases.

Perceived Disadvantages: There are several areas, at first glance, that may appear to be a disadvantage, but when viewed as part of the bigger picture may even result in a favorable **situation.**

Model managers have less control: Individual model managers will have less control than previously because models and data must be coordinated now. This coordination may make it more difficult for a given model manager to immediately make model improvements, but at the same time be an advantage since the entire toolset should become more integrated rather than having a set of independent, stand-alone models.

Program managers may have less control: With a common toolset in place as an expected starting point for most studies, many program managers will not have as much freedom to construct a study to their individual choosing. However, this may also be an advantage from the perspective of senior government leaders who want oversight and some consistency across studies and programs.

Companies pay more user fees: Companies will probably pay more in users fees, because there will be more users groups. However, this cost should be more than offset by the amount of savings from no longer having to build company-unique databases for every model in a company's portfolio.

Companies cannot use their own models: This statement is partially true but is not necessarily a disadvantage. For analysis cases where either the Pedigreed models or Pedigreed databases are not appropriately sensitive or are otherwise inappropriate, contractors are free to make their case. In such an instance, there is no prohibition from "doing the study right." However, where it is not the case, the advantages gained from requiring their use outweigh the simple desire by the contractors to use their own tools.

PROTECTIVE SAFEGUARDS

One of the principal objectives of the Database Standardization Committee was to recommend safeguards that provide protection to avoid misuse of certified data while expanding the availability of this data to the broader community. It became immediately apparent that preventing misuse of certified data must also consider the models and model assumptions as well as the data. Anyone utilizing "Certified Data" can still fix or bias the outcome of any study by manipulating the model, modeling assumptions, and/or inputs not covered by the certified data and claim the results are valid because they are based on the certified data. As a result, safeguards that considered this larger problem of data misuse are also addressed.

Recommended safeguards can be categorized in terms of answers to the following questions:

1. Should there be a study certification process applied to the results of studies utilizing certified data to discourage misuse?

2. If so, which studies should be certified and by whom?

3. Should denying some users access to certified data prevent misuse? Under what conditions should potential users should be denied access to data?

4. Who decides whether misuse has occurred?

5. What if any punitive actions can or should be taken against an organization that has been judged as misusing the database?

6. What different levels of safeguards need to be applied to different kinds of data?

    - Proprietary versus Non-Proprietary
    - Model-Dependent versus Non-Model-Dependent

7. What measures can be taken to expand the availability of certified data to the community?

The paragraphs below address each of these questions separately, providing the committee's recommendations on each subject.

Should there be a study certification process applied to the results of studies utilizing certified data to discourage misuse?

The committee believes the answer to this question is "yes." However, they also recognize a significant commitment in time and dollars may be required to provide this service. Successful implementation of this process will require certification level or ratings be established and consistently applied. One hypothetical scheme would identify certification levels as follows:

Level 0: Data utilized is inconsistent with certified database.

Level 1: Data utilized is consistent with certified database but results are not deemed credible due to unapproved models or modeling assumptions.

Level 2: Data, models, and modeling assumptions used are consistent with approved models and certified database. Results are deemed credible.

With this hypothetical plan, a three-element Study Rating Scheme of [D, M, R] can also be used to represent the above certification process, where:

1. "D" indicates whether data utilized is consistent with certified database. Yes = 1, No = 0.

2. "M" indicates whether or not models and modeling assumptions are consistent with approved models. Yes = 1, No = 0, Not Rated = N.

3. "R" indicates whether or not results are deemed logical and credible. Yes = 1, No = 0, Not Rated = N.

Under these schemes, a Level 1 Certification or a Study Rating of [1,0,0] or [1,1,0] would indicate potential misuse of certified data.

What studies should be certified and by whom?

Any study that relies heavily on the kind of data available in the certified database, whether performed by a government agency or contractor, should be able to obtain study certification upon request. Initially, only government-funded studies above a specified dollar amount should be considered.

The Government Program Manager should grant actual certification for contractor studies. Air Force Studies and Analysis (or CNA, TRADOC) should take on responsibility for certifying studies performed by other service and DoD agencies. Outside government agencies can also certify or not certify study results if they

disagree with program office certification. Opposing views must be backed up with clear rationale as to why a study failed to meet criteria for certification.

Should denying some users access to certified data or should potential users be denied access to this data under certain conditions prevent misuse?

Since the objective is to maximize availability while safeguarding data, the committee recommends the certification process itself is used to provide safeguards rather than denying users access. However, it is recommended that users requesting data must sign an agreement to submit their study results on the request of the appropriate government agency for certification. Failure to sign such an agreement should be the only reasonable grounds for denying access.

Who determines whether misuse has occurred?

The certification process identifies suspected misuse. Anyone can point to a lack of certification or poor certification rating as basis for concern.

What if any punitive actions can or should be taken against an organization that has been judged as misusing the database?

No punitive actions need be taken for misusing the data. Failure to seek and/or achieve study certification should be used as key indicator in assessing the credibility of the results, just as a contractor's failure to use an industry standard or program-approved model would be. A consistent track record on the part of a contractor to seek and achieve certification should be considered a competitive advantage and integrated into past performance measures for proposal evaluation.

What different levels of safeguards need to be applied to different kinds of data?

Proprietary or competition sensitive data must be identified as such and excluded from files accessible by non-government contractors. Government agencies can have access to all data. Some data may need restrictions and/or caveats as to its allowable use based on the models and other assumptions associated with its creation. For classified data, existing DoD directives are adequate. This includes "need to know" and "special access" procedures.

What measures can be taken to expand the availability of certified data to the community?

- The government may need to expend a significant effort to create non-proprietary data representing proprietary system capabilities for other contractors to use.
- Data having a "For Program X Use Only" caveat will not be allowed as part of the certified database.

- Approved users should be able to access needed data files via White Knight or other secure modem connection at any time, rather than requiring approval for each request. Access to these files must be consistent with the security level and whether or not the information is proprietary.
- A simple, easy to use process must be in place that will allow contractors and government agencies to add to the certified database.

PRIORITIZATION CRITERIA

The intent of this committee has been to concentrate on standardized databases and not on models. However, models and databases must be considered together when establishing prioritization criteria. After all, a model is almost useless without a database, and a database cannot provide an analyst with much of a study without a model or two.

The ideal situation would be to have one central database that could be mapped into model specific databases. Then only the central database would have to be maintained. Changes to the model could impact the mapping algorithms, but generally should not impact the central database. However, this ideal situation is not reality today. The following paragraphs suggest criteria to prioritize the order in which data for USAF models should be developed.

Criterion 1. Provide data for the hierarchy (pyramid) of approved models.

The highest priority for a standardized database should go to approved government or government sponsored models. The so-called "Pedigree Database" is intended to support a hierarchy of models: engineering level; engagement level; mission level and campaign level models.

A major consideration should be the database breadth. Do all services contribute to and use the database or is it primarily of interest to just one service? The database should be broadly available throughout the government and to contractors. If the database contains any program sensitive data, then it should also contain substitute data that can be shared outside the program.

At the present time, the Air Force is supporting a limited set of "Toolkit" models to be used for official Air Force AOA.

Criterion 2. Develop data for approximately one campaign model, two mission models, four engagement models, and eight engineering level models.

The overall objective is to reduce the M&S costs. Reusable databases and fewer models can help lower these costs. Since it is desirable to reduce the total number of models used, standardized databases should be developed for a balanced set of models. Each level of model must be represented. As the shape of the pyramid suggests, there must be more lower level models than campaign models. Therefore, a

few lower level models (e.g., three engagement and two mission level models) and only one campaign level model should be given priority for standardized data development. Level of detail should be kept consistent whenever possible.

Criterion 3. Provide engineering and engagement level data before mission and campaign data.

Since the output from the lower level models is used in the development of data for the higher level models, a standardized database should be developed for the lower level models first, then the higher level models.

Criterion 4. Provide well documented databases.

Does the data have an audit trail to validated source data? Was the data derived using sound analytical methods? Completeness of the database should be considered. How much work would be required to complete the database?

Criterion 5. Keep the databases current for the selected (approved) models. This takes priority over developing standardized databases for other models.

As the models are updated, the standardized databases supporting those models should be kept current as well, even at the expense of not developing new standardized databases for other models. Such a focused effort helps lower costs.

In summary, the criteria listed below should be used to prioritize models for which standardized databases should be developed. For example, applying those criteria today would lead to the selection of databases that support models such as THUNDER at the campaign level. The databases should support the Pedigreed models. The databases should be documented and have an audit trail and consistent level of detail. The databases should be capable of being mapped into databases for the J-series of models when they come online.

The criteria listed above lead to the following databases that should be developed initially:

1. Campaign Level: THUNDER
2. Mission Level: Choose 1 or 2 from: SUPPRESSOR, SWEG, EADSIM
3. Engagement Level: ESAMS, RADGUN, BRAWLER, MOSAIC
4. Engineering Level: FASTGEN/COVART, ALARM, TRAP, BLUE MAX, etc.

IMPLEMENTATION ACTIONS

Throughout the investigation of data, the NDIA Data Base Standardization Committee learned of various tools that were acceptable and available to analytic communities in large portions of industry and the Air Force. Likewise, the committee discovered there was much groundwork required if these analytic communities expected to fully utilize DoD-managed "coherent, consistent, documented data." On the other hand, the committee observed pockets of "work in progress" by the various services both singular and joint throughout the DoD. As such, the committee felt any implementation actions should embrace new initiatives and capture the innovative work in progress.

As reflected throughout this report, the committee research concentrated mainly on "**what**" should be implemented rather than "**how.**" Consequently, the implementation action recommendations are succinct with much of the follow-on mechanics falling on the shoulders of our recommended DoD sponsor.

Fundamentally, implementation should foster an architecture that would permit all users (both government and industry) to share managed data—coherent, consistent, documented data. Additionally, this architecture must service today's legacy models along with providing the database underpinning for the next generation of joint models currently under development. Implementation, in the simplest form, is viewed as a two-step process with AF-wide acceptance as the cornerstone.

Step 1 tackles the fundamentals:

1. Writing a policy for "Pedigreed" database use and management.

2. Establishing a standardized legacy model toolkit and limiting the supported models to those with the greatest usefulness and credibility.

3. Allocating funding to keep the implementation process and continued development of Pedigree databases healthy.

Step 2 is the insurance policy for Step 1. Establish AF/XOC as the implementation sponsor and the Pedigree process owner. As such, it is suggested that AF/XOC establish a Modeling and Simulation Data Management Integrated Project Team (IPT) to explore actions addressed in Step 1 and the validity of the investigations made throughout this report.

This suggested implementation action plan focuses on the Air Force and AF/XOC in particular. However, it is important to emphasize the awareness and initiatives surfacing throughout the DoD M&S communities that touch on the needs for managed (consistent, coherent, and documented) data. In this regard, if the Pedigree process advocated in this report is to survive in the long term, it must be concurrently accepted not only by the Air Force but also through actions underway by the various services, Tri-Service teams, and DMSO.

SECTION 4 — CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

The ad hoc committee verified that, in general, databases were not maintained and updated for reuse. It concluded that adopting a process that would support a preferred set of standardized models was desirable. The group agreed that standardizing databases for reuse requires the development of consistent, coherent, and documented data. The key to this Pedigree process is the acceptance and application of data standards by all data sources supporting the process. The committee concluded that safeguards would have to be implemented to prevent misuse of the data. Finally, the group agreed that the benefits far outweigh the actions required for implementation of the process.

RECOMMENDATIONS

The committee recommends:

- ☑ **Write a policy for Pedigreed database use and management.**
- ☑ **Once the policy is in place, prioritize the models used most frequently and apply the Pedigree process to the databases for those models.**
- ☑ **Provide sufficient funding to initiate and sustain the process. Part of the process must include certification to provide credibility and prevent misuse.**
- ☑ **Establish AF/XOC (HQ USAF, Director, Command and Control, Deputy Chief of Staff/Air and Space Operations) as the implementation sponsor and Pedigree process owner.**
- ☑ **Following policy implementation, NDIA should investigate the technical aspects of the implementation.**
- ☑ **Make this report available through NDIA and have NDIA assist in placing it "on-line" with the Defense Technical Information Committee (DTIC).**

APPENDIX A — Briefing to HQ U.S. Air Force, Command and Control, Deputy Chief of Staff, Air and Space Operations, Department of the Navy, Office of the CNO (N81) and the National Defense Industrial Association

Slide 1

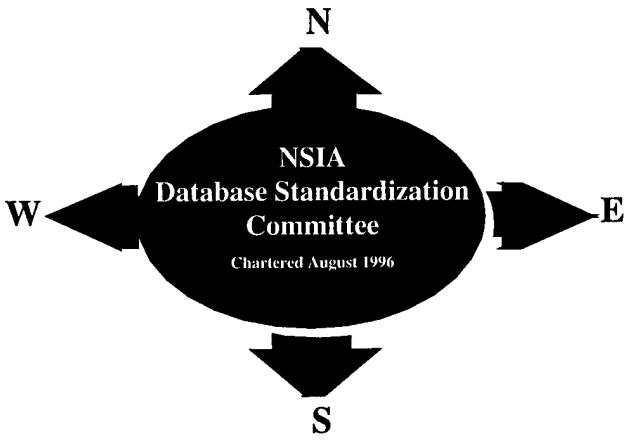
NSIA
Database Standardization Committee
Aug 96 - May 97

Vinton Cline
Chairman

Scott Saunders
Presenter for LTG SKIBBIE

Slide 2

Heading Check



AGENDA

PURPOSE

WHO WE ARE

THE PROBLEM

PROPOSAL

BOTTOM LINE

NEXT STEP

Slide 3

Adhoc Committee
Background & Approach

NSIA Database Standardization Committee (NDSC) Origin
-- NSIA Opportunity to explore on-going "Pedigree Database" initiative

PHASE I Organizing

PHASE II Investigating

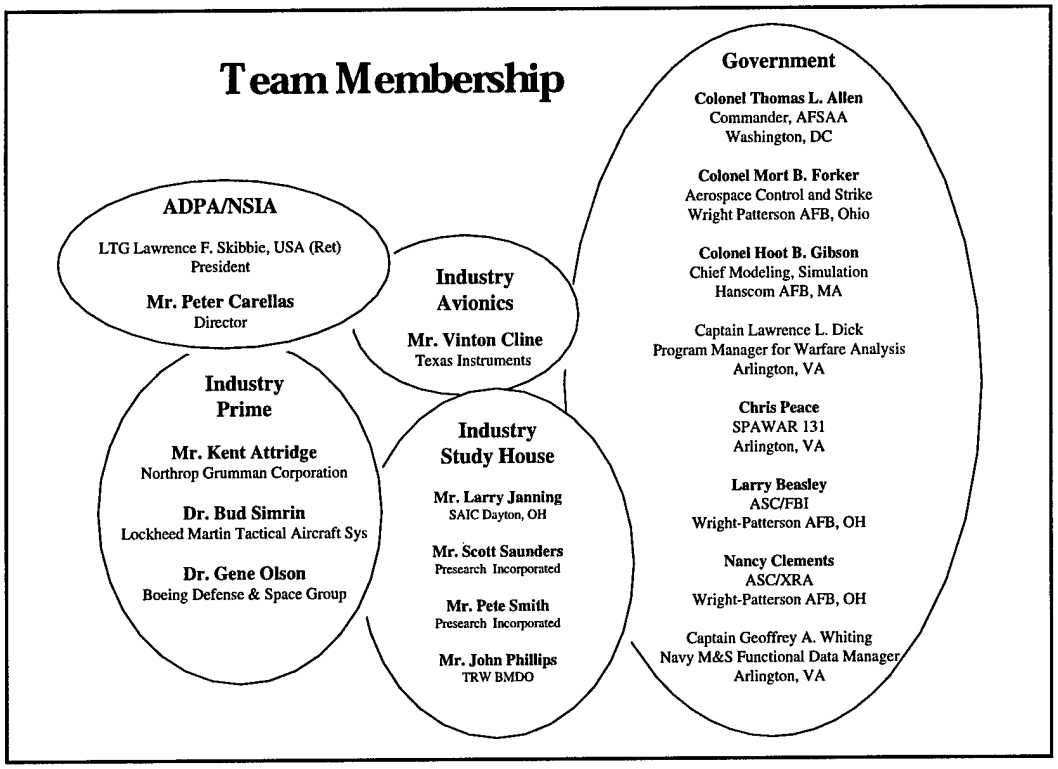
- Accomplished Investigative Analysis
- Expand Objectives & Alternatives
- Made Observations & Recommendations

PHASE III Heading Check

PHASE IV Update/Incorporate "Heading Check" Feedback

PHASE V Report & Brief to Government / Industry

Slide 4



Slide 5

Problem Statement

UNLIKE MODEL MANAGEMENT, NO COMMONLY AGREED UPON DIRECTION EXISTS FOR COHERENT, CONSISTENT, DOCUMENTED (C_ C_D) SYSTEM DATA

QUICK-REACTION STUDY RESPONSE IS DIFFICULT WHEN LARGE DATABASES MUST BE REBUILT

TOO MUCH STUDY TIME AND MONEY WASTED ON UP-FRONT DATABASE EFFORTS

INDUSTRY/GOVERNMENT ESTIMATES: 30-80% OF STUDY COSTS CAN BE SAVED THROUGH A STANDARD REFERENCED DATABASE PROCESS

Slide 6

Committee Action Items

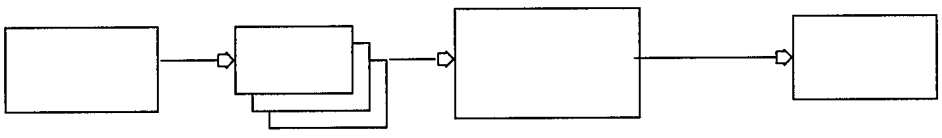
1. IDENTIFIED DATA SOURCES
2. RECOMMENDED STANDARDS AND PROCESSES
3. LOOKED AT DATA SHARING METHODS
4. IDENTIFIED ADVANTAGES AND DISADVANTAGES
5. LOOKED AT WAYS TO AVOID MISUSE
6. PRIORITIZED MODELS NEEDING DATABASES
7. PROPOSED AN IMPLEMENTATION PLAN

ACTION ITEMS LED TO A BETTER UNDERSTANDING OF EXISTING DATABASE ISSUES AND PROGRAMS, AND IDENTIFIED A MAJOR VOID IN CURRENT DATABASE SUPPORT TO THE STUDY PROCESS

Slide 7

Current Study Process

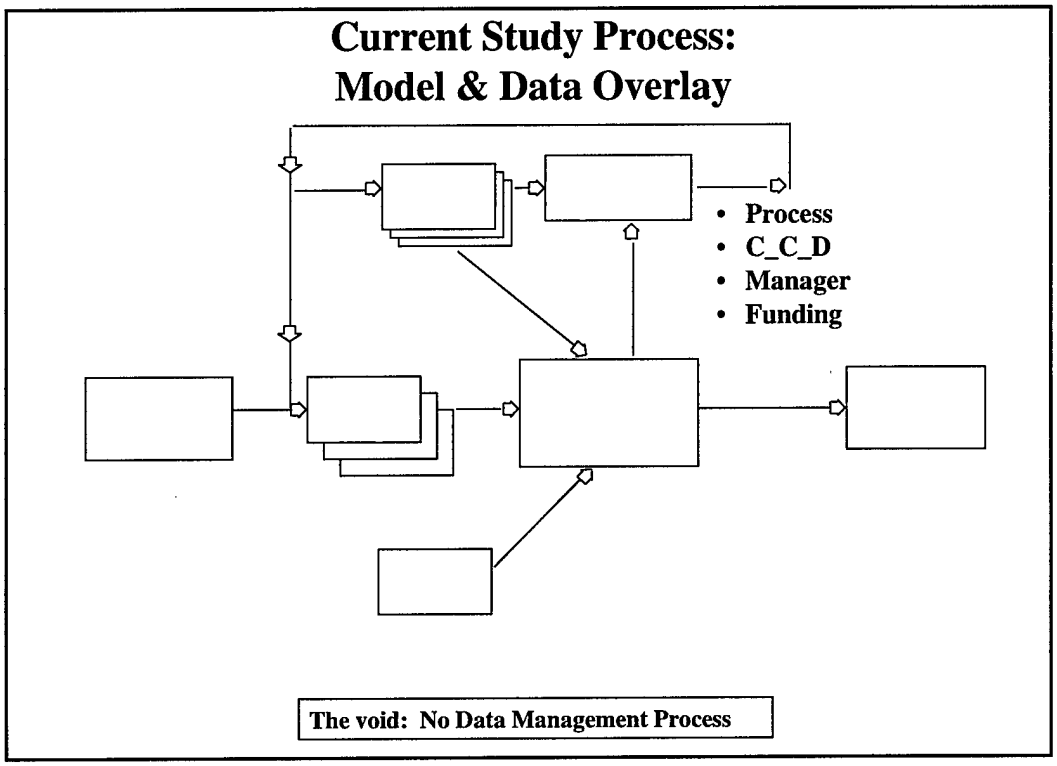
To illustrate the void our committee discovered



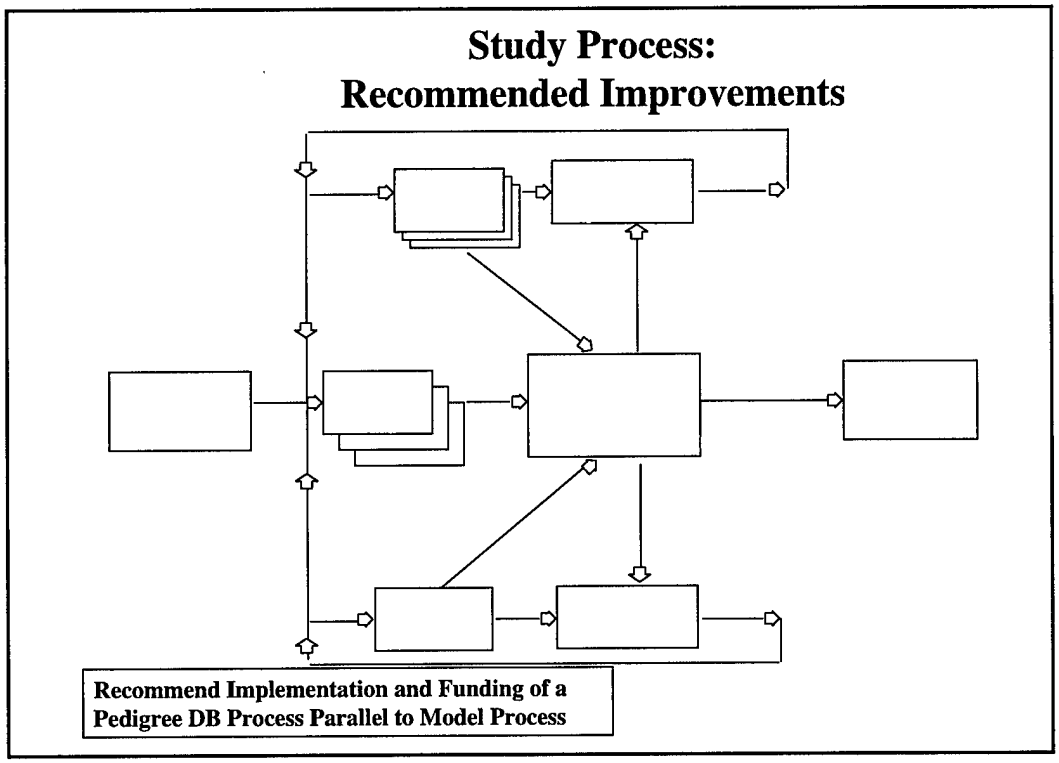
- **Congress**
- **OSD**
- **JCS**
- **Services**
- **Industry**
- **Programs**
- **Warfighters**

- **AOA**
- **MNS**
- **MAA**
- **Etc**

Slide 8



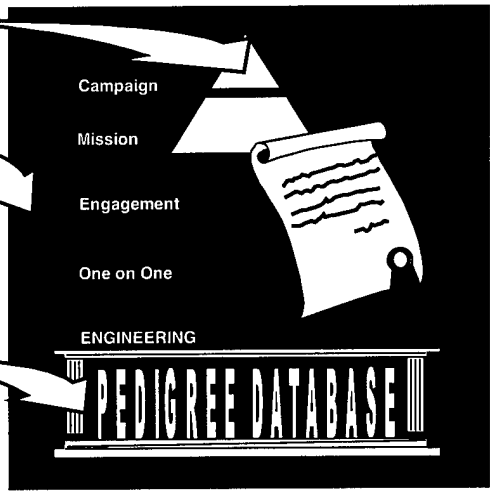
Slide 9



Slide 10

Pedigree Elements

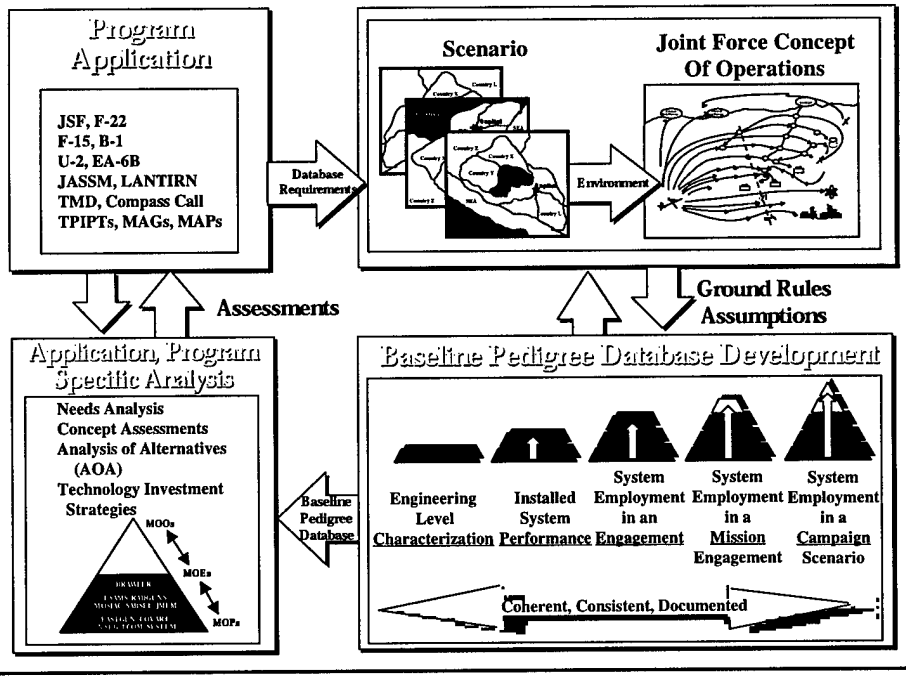
1. Standard Hierarchy of Constructive ***Models***
2. Integration ***Processes*** for development of consistent data throughout Hierarchy of Models
3. ***Databases*** that have been documented and have an audit trail specifying data sources



Pedigree process satisfies the NSIA Coherent, Consistent, Documented (C_C_D) Requirement

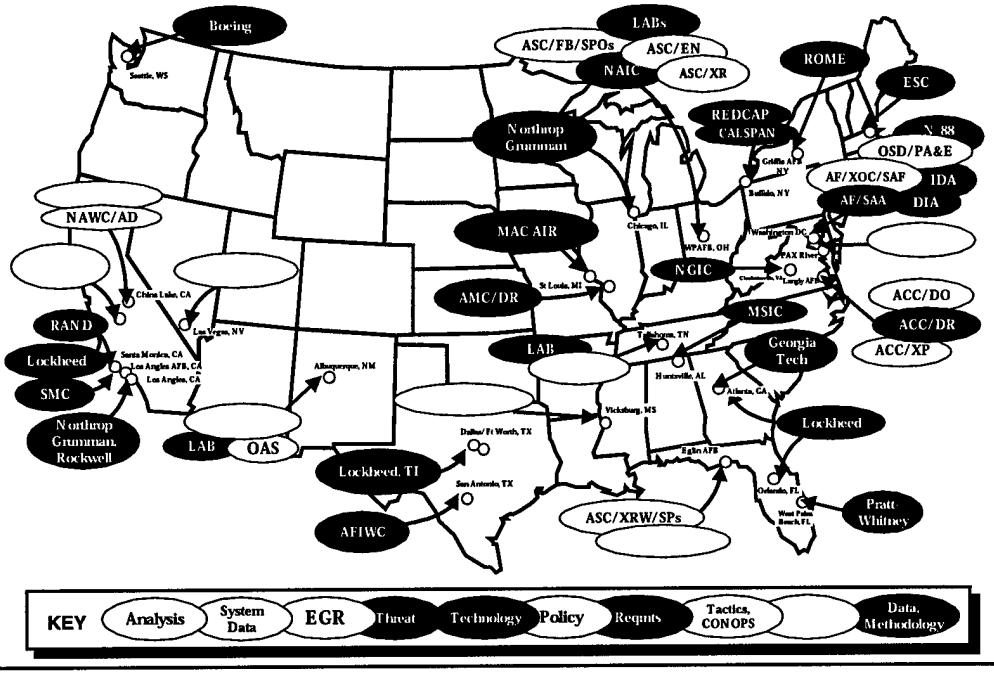
Slide 11

Baseline Pedigree Database Development



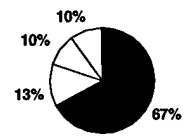
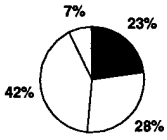
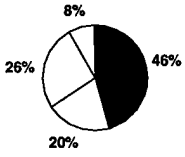
Slide 12

Current Pedigree Database Interactions



Slide 13

Pedigree Database Status

	Data Consistency	Data Quality	Overall Data Status
FY 96			

Data Consistency: ***Consistent, traceable (audit trail), derived data from One-on-one Level through Campaign Level and across functional Areas***
Current Focus: Survivability

Data Consistency

	No Lower Level Anlaysis Audit Trail
	Known Lower Level Audit Trail with obsolete Methodology
	Known Lower Level Audit Trail, Latest Methodology
	Surrogated

Data Quality - ***Derived from sound analytical methods or from an accredited source***

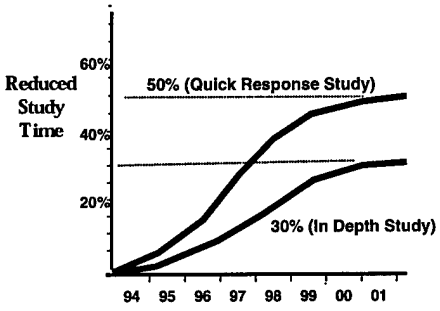
Data Quality

	No Data, Deficiency
	Unknown Origin, Obsolete, Slight Deficiencies, Functionally not verified
	Documented Source, Current, Sufficient, Functionally Verified
	Surrogated

Slide 14

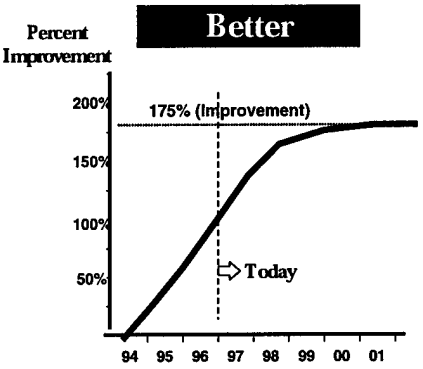
Estimated Benefits[1]

Faster



- Less Lead Time
- Shorter Learning Curve
- Reduces Government Adjudication
- Faster Documentation

Better



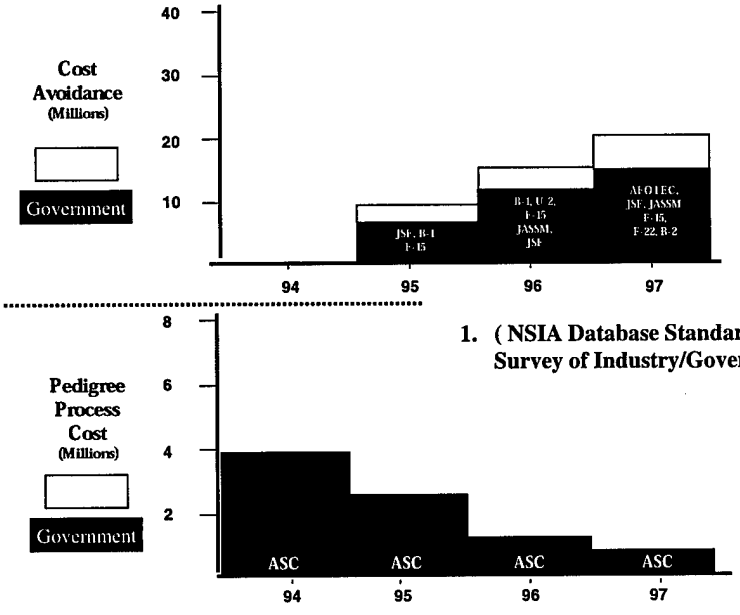
- Audit Trail
- Apples to Apples Comparisons
- Data Consistency
- Documentation
- Configuration Management
- High Confidence in Results

(1) NSIA Database Standardization Committee Research

Slide 15

Estimated Benefits[1] (Cont'd)

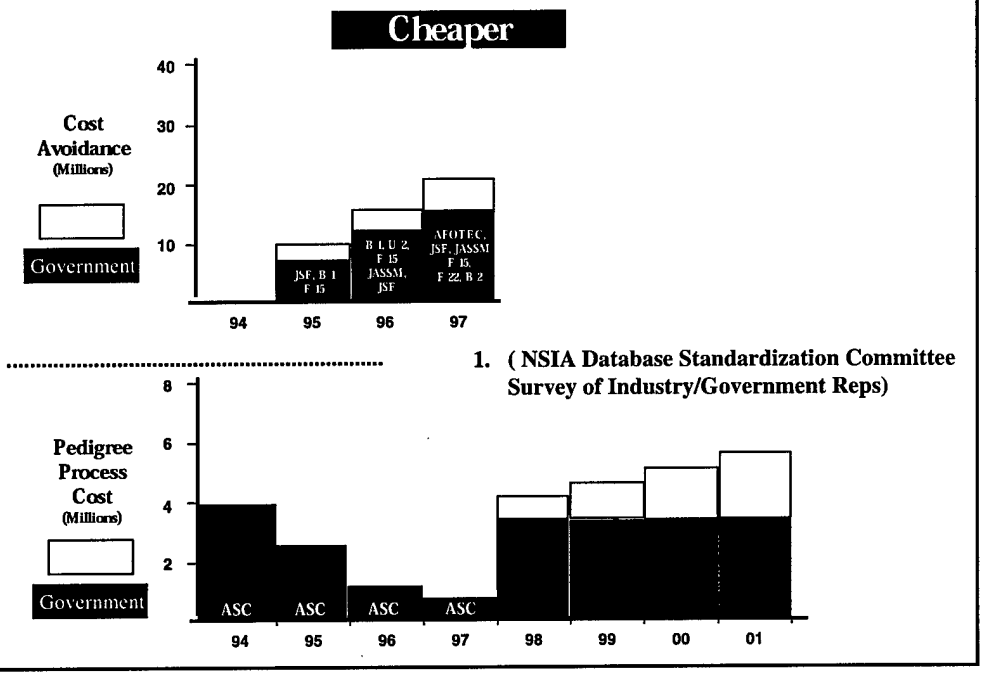
Cheaper



1. (NSIA Database Standardization Committee Survey of Industry/Government Reps)

Slide 16

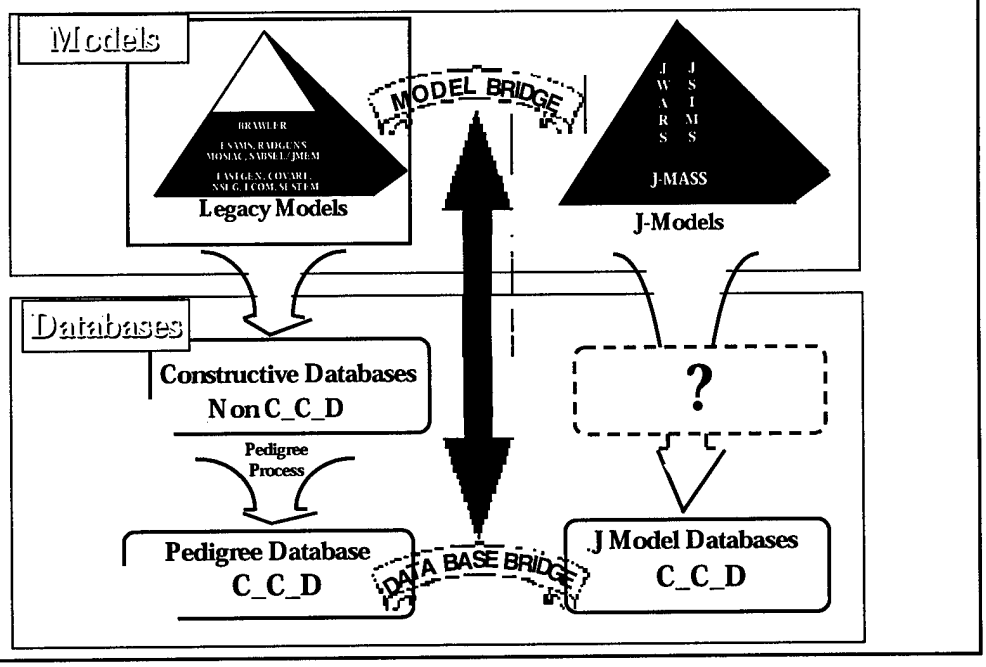
Estimated Benefits[1] (Cont'd)



1. (NSIA Database Standardization Committee Survey of Industry/Government Reps)

Slide 17

Bridging Today and Tomorrow



Slide 18

Heading Check Status

BRIEFED LTG SKIBBIE (ADPA / NSAI SPONSOR)

ON AN INDIVIDUAL BASIS, SPONSOR PRESENTATION TO:

INDUSTRY

DoD (SERVICES: OSD, JOINT STAFF)

ASC HAS OFFERED RESOURCES TO PRODUCE VIDEO WHICH WILL BE COORDINATED THROUGH ADPA/NSIA

DRAFT RECOMMENDATIONS FOR AF/XOC CONSIDERATION

Slide 19

Draft Recommendations (AF/XOC)

ESTABLISH XOC-LED IPT TO REVIEW IMPLEMENTATION OF THE FOLLOWING:

AF-WIDE IMPLEMENTATION OF PROCESS

- WRITE POLICY FOR PEDIGREE DATABASE USE AND MANAGEMENT
- ALLOCATE FUNDING NOW TO PRIORITIZE REQUIREMENTS PROCESS
- ESTABLISH PEDIGREE USERS GROUP/CONFIG MANAGEMENT PROCESS (XOC LEAD?)
- CREATE LIBRARY (FORMAT, FEES, SECURITY, DISTRIBUTION, ETC) (XOC LEAD?)

SELECT A PROCESS OWNER FOR PEDIGREE DATABASE (XOC?)

- ASSIGN ACTION OFFICER
- TASK AFMC MSIO, AFSAA, AFAMS TO SUPPORT INITIATIVE

SPONSOR BRIEFINGS: SAF/AQ, AF/TE, AF/XO/XOR, ACC/DR, AMC/DR, AFMC/ST

INTEGRATE PEDIGREE DATABASES INTO AF M&S VISION

SUPPORT BROADER RELEASE OF PEDIGREE TO INDUSTRY, CONSISTENT WITH POLICY

EXPAND PEDIGREE INITIATIVE TO INCLUDE OTHER SERVICES AND DoD

GOVT.. HOST WORKING GROUP(S) TO DRAFT TRANSITION ROADMAPS FROM LEGACY DATABASES TO JTRIAD DATABASE SUPPORT

APPENDIX B — AF/XOC Response to NDIA Database Standardization Committee Briefing

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS UNITED STATES AIR FORCE
WASHINGTON, DC

05 DEC 1997

HQ USAF/XOC
1480 Air Force Pentagon
Washington DC 20330-1480

Lt Gen Lawrence F. Skibbie, Ret.
National Defense Industrial Association
2111 Wilson Boulevard, Suite 400
Arlington, Virginia 22201-5000

Dear General Skibbie

This past May, you arranged for Messrs Vinton Cline and Scott Saunders to brief the National Security Industrial Association Database Standardization committee program to Major General Charles R. Henderson, my predecessor as Director of Command and Control (AF/XOC). That briefing included a recommendation for AF/XOC to establish an integrated process team (IPT) to incorporate Pedigree database processes in Air Force programs.

We would like to update you on our progress. As you recommended, my staff took the lead and investigated the need for such a program across the Air Force. The results of the investigation were positive, and I tasked them to develop a plan to extend the Pedigree database process Air Force-wide and find ways to have the process adopted throughout the Department of Defense. We have found that not only does Pedigree Database procedures provide a template to help us assure quality data, but the procedures also provide a means for ensuring coherent, consistent data throughout aggregation levels from engineering level models through campaign level models. We have included Pedigree as one of the focus areas of our recently formed Modeling and Simulation Data Management IPT so as to both improve the Pedigree process and expand its impact within the AF. Additionally, we are attempting to influence the DoD and other Services approach to authoritative data representation and quality (VV&C) by making the Pedigree process a key element of our contribution to the Tri-Service VV&C Tiger Team. Additionally, we have suggested to the DMSO point of contact for the Tri-Service VV&C Tiger Team, Mr. Bill Dunn, that the results from your Database Standardization Committee might be very helpful to that committee.

I anticipate the Pedigree database process will facilitate a smooth transition and help us to achieve the Air Force vision of a Joint Synthetic Battlespace supporting better decisions and warfighting skills. Additionally, we hope that other Services and Dod agencies will adopt these procedures leading to reusable, quality data. If you or your staff have further comments or questions, please contact Colonel Lynne Thompson. His telephone number is (202) 761-5340 ext 101.

Sincerely

[redacted] USAF
[redacted]
DCS/Air and Space Operations

cc:
Mr Vinton Cline
Mr Scott Saunders